

STATISTICS



Principles & Methods

RICHARD A. JOHNSON | GOURI K. BHATTACHARYYA

6TH EDITION



This online teaching and learning environment integrates the entire digital textbook with the most effective instructor and student resources to fit every learning style.

With WileyPLUS:

• Students achieve concept mastery in a rich, structured environment that's available 24/7

• Instructors personalize and manage their course more effectively with assessment, assignments, grade tracking, and more



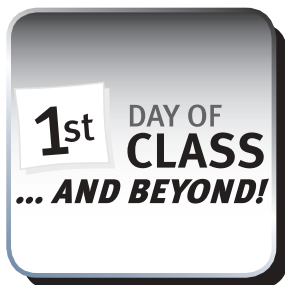
- manage time better
- study smarter
- save money

From multiple study paths, to self-assessment, to a wealth of interactive visual and audio resources, *WileyPLUS* gives you everything you need to personalize the teaching and learning experience.

» Find out how to MAKE IT YOURS »

www.wileyplus.com

ALL THE HELP, RESOURCES, AND PERSONAL SUPPORT YOU AND YOUR STUDENTS NEED!



2-Minute Tutorials and all of the resources you & your students need to get started
www.wileyplus.com/firstday



Student support from an experienced student user
Ask your local representative for details!



Collaborate with your colleagues, find a mentor, attend virtual and live events, and view resources
www.WhereFacultyConnect.com



Pre-loaded, ready-to-use assignments and presentations
www.wiley.com/college/quickstart



Technical Support 24/7
FAQs, online chat, and phone support
www.wileyplus.com/support



Your WileyPLUS Account Manager
Training and implementation support
www.wileyplus.com/accountmanager

Statistics

Principles and Methods

SIXTH EDITION

Richard A. Johnson

University of Wisconsin at Madison

Gouri K. Bhattacharyya



John Wiley & Sons, Inc.

Vice President & Executive Publisher
Project Editor
Senior Development Editor
Production Manager
Senior Production Editor
Marketing Manager
Creative Director
Design Director
Production Management Services
Photo Editor
Editorial Assistant
Media Editor
Cover Photo Credit

Cover Designer

Laurie Rosatone
Ellen Keohane
Anne Scanlan-Rohrer
Dorothy Sinclair
Valerie A. Vargas
Sarah Davis
Harry Nolan
Jeof Vita
mb editorial services
Sheena Goldstein
Beth Pearson
Melissa Edwards
Gallo Images-Hein von
Horsten/Getty Images, Inc.
Celia Wiley

This book was set in 10/12 Berling by Laserwords Private Limited, India and printed and bound by RR Donnelley-Crawsfordville. The cover was printed by RR Donnelley-Crawsfordville.

Copyright © 2010, 2006 John Wiley & Sons, Inc. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc. 222 Rosewood Drive, Danvers, MA 01923, website www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774, (201)748-6011, fax (201)748-6008, website <http://www.wiley.com/go/permissions>.

Evaluation copies are provided to qualified academics and professionals for review purposes only, for use in their courses during the next academic year. These copies are licensed and may not be sold or transferred to a third party. Upon completion of the review period, please return the evaluation copy to Wiley. Return instructions and a free of charge return shipping label are available at www.wiley.com/go/returnlabel. Outside of the United States, please contact your local representative.

ISBN-13 978-0-470-40927-5

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Preface

THE NATURE OF THE BOOK

Conclusions, decisions, and actions that are data driven predominate in today's world. Statistics—the subject of data analysis and data-based reasoning—is necessarily playing a vital role in virtually all professions. Some familiarity with this subject is now an essential component of any college education. Yet, pressures to accommodate a growing list of academic requirements often necessitate that this exposure be brief. Keeping these conditions in mind, we have written this book to provide students with a first exposure to the powerful ideas of modern statistics. It presents the key statistical concepts and the most commonly applied methods of statistical analysis. Moreover, to keep it accessible to freshmen and sophomores from a wide range of disciplines, we have avoided mathematical derivations. They usually pose a stumbling block to learning the essentials in a short period of time.

This book is intended for students who do not have a strong background in mathematics but seek to learn the basic ideas of statistics and their application in a variety of practical settings. The core material of this book is common to almost all first courses in statistics and is designed to be covered well within a one-semester course in introductory statistics for freshmen–seniors. It is supplemented with some additional special-topics chapters.

ORIENTATION

The topics treated in this text are, by and large, the ones typically covered in an introductory statistics course. They span three major areas: (i) descriptive statistics, which deals with summarization and description of data; (ii) ideas of probability and an understanding of the manner in which sample-to-sample variation influences our conclusions; and (iii) a collection of statistical methods for analyzing the types of data that are of common occurrence. However, it is the treatment of these topics that makes the text distinctive. Throughout, we have endeavored to give clear and concise explanations of the concepts and important statistical terminology and methods. By means of good motivation, sound explanations, and an abundance of illustrations given in a real-world context, it emphasizes more than just a superficial understanding.

Each statistical concept or method is motivated by setting out its goal and then focusing on an example to further elaborate important aspects and to illustrate its application. The subsequent discussion is not only limited to showing how a method works but includes an explanation of the why. Even without recourse to mathematics, we are able to make the reader aware of possible pitfalls in the statistical analysis. Students can gain a proper appreciation of statistics only when they are provided with a careful explanation of the underlying logic. Without this understanding, a learning of elementary statistics is bound to be rote and transient.

When describing the various methods of statistical analysis, the reader is continually reminded that the validity of a statistical inference is contingent upon certain model assumptions. Misleading conclusions may result when these assumptions are violated. We feel that the teaching of statistics, even at an introductory level, should not be limited to the prescription of methods. Students should be encouraged to develop a critical attitude in applying the methods and to be cautious when interpreting the results. This attitude is especially important in the study of relationship among variables, which is perhaps the most widely used (and also abused) area of statistics. In addition to discussing inference procedures in this context, we have particularly stressed critical examination of the model assumptions and careful interpretation of the conclusions.

SPECIAL FEATURES

1. **Crucial elements are boxed** to highlight important concepts and methods. These boxes provide an ongoing summary of the important items essential for learning statistics. At the end of each chapter, all of its **key ideas and formulas** are summarized.
2. **A rich collection of examples and exercises** is included. These are drawn from a large variety of **real-life settings**. In fact, many data sets stem from genuine experiments, surveys, or reports.
3. **Exercises** are provided at the end of **each major section**. These provide the reader with the opportunity to practice the ideas just learned. Occasionally, they supplement some points raised in the text. A larger collection of exercises appears at the **end of a chapter**. The starred problems are relatively difficult and suited to the more mathematically competent student.
4. **Using Statistics Wisely**, a feature at the end of each chapter, provides important guidelines for the appropriate use of the statistical procedures presented in the chapter.
5. **Statistics in Context** sections, in four of the beginning chapters, each describe an important statistical application where a statistical approach to understanding variation is vital. These extended examples reveal, early on in the course, the value of understanding the subject of statistics.
6. **P-values** are emphasized in examples concerning tests of hypotheses. Graphs giving the relevant normal or t density curve, rejection region, and P -value are presented.

7. **Regression analysis** is a primary statistical technique so we provide a more thorough coverage of the topic than is usual at this level. The basics of regression are introduced in Chapter 11, whereas Chapter 12 stretches the discussion to several issues of practical importance. These include methods of **model checking**, handling nonlinear relations, and multiple regression analysis. Complex formulas and calculations are judiciously replaced by computer output so the main ideas can be learned and appreciated with a minimum of stress.
8. **Integrated Technology**, at the end of most chapters, details the steps for using MINITAB, EXCEL,¹ and TI-84 calculator. With this presentation available, with few exceptions, only computer output is needed in the text. Software packages remove much of the drudgery of hand calculation and they allow students to work with larger data sets where patterns are more pronounced. Some computer exercises are included in all chapters where relevant.
9. **Convenient Electronic Data Bank** at the end of the book contains a substantial collection of data. These data sets, together with numerous others throughout the book, allow for considerable flexibility in the choice between concept-orientated and applications-orientated exercises. The Data Bank and the other larger data sets are available for download on the accompanying Web site located at www.wiley.com/college/johnson.
10. **Technical Appendix A** presents a few statistical facts of a mathematical nature. These are separated from the main text so that they can be left out if the instructor so desires.

ABOUT THE SIXTH EDITION

The sixth edition of *STATISTICS—Principles and Methods* maintains the objectives and level of presentation of the earlier editions. The goals are the developing (i) of an understanding of the reasonings by which findings from sample data can be extended to general conclusions and (ii) a familiarity with some basic statistical methods. There are numerous data sets and computer outputs which give an appreciation of the role of the computer in modern data analysis.

Clear and concise explanations introduce the concepts and important statistical terminology and methods. Real-life settings are used to motivate the statistical ideas and well organized discussions proceed to cover statistical methods with heavy emphasis on examples. The sixth edition enhances these special features. The major improvements are:

Bayes' Theorem. A new section is added to Chapter 4 to highlight the reasoning underlying Bayes's theorem and to present applications.

Approximate t . A new subsection is added to Chapter 7, which describes the approximate two sample t statistic that is now pervasive in statistical software programs. For normal distributions, with unequal variances, this has become the preferred approach.

¹Commands and the worksheets with data sets pertain to EXCEL 2003.

New Examples. A substantial number of new examples are included, especially in the core chapters, Chapter 11 on regression, and Chapter 13 on contingency tables.

More Data-Based Exercises. Most of the new exercises are keyed to new data-based examples in the text. New data are also presented in the exercises. Other new exercises are based on the credit card use and opinion data that are added to the data bank.

New Exercises. Numerous new exercises provide practice on understanding the concepts and others address computations. These new exercises, which augment the already rich collection, are placed in real-life settings to help promote a greater appreciation of the wide span of applicability of statistical methods.

ORGANIZATION

This book is organized into fifteen chapters, an optional technical appendix (Appendix A), and a collection of tables (Appendix B). Although designed for a one-semester or a two-quarter course, it is enriched with ample additional material to allow the instructor some choices of topics. Beyond Chapter 1, which sets the theme of statistics and distinguishes population and sample, the subject matter could be classified as follows:

Topic	Chapter
Descriptive study of data	2, 3
Probability and distributions	4, 5, 6
Sampling variability	7
Core ideas and methods of statistical inference	8, 9, 10
Special topics of statistical inference	11, 12, 13, 14, 15

We regard Chapters 1 to 10 as constituting the core material of an introductory statistics course, with the exception of the starred sections in Chapter 6. Although this material is just about enough for a one-semester course, many instructors may wish to eliminate some sections in order to cover the basics of regression analysis in Chapter 11. This is most conveniently done by initially skipping Chapter 3 and then taking up only those portions that are linked to Chapter 11. Also, instead of a thorough coverage of probability that is provided in Chapter 4, the later sections of that chapter may receive a lighter coverage.

SUPPLEMENTS

Instructor's Solution Manual. (ISBN 978-0-470-53519-6) This manual contains complete solutions to all exercises.

Test Bank. (Available on the accompanying website: www.wiley.com/college/johnson) Contains a large number of additional questions for each chapter.

Student Solutions Manual. (ISBN 978-0-470-53521-9) This manual contains complete solutions to all odd-numbered exercises.

Electronic Data Bank. (Available on the accompanying website: www.wiley.com/college/johnson) Contains interesting data sets used in the text but that can be used to perform additional analyses with statistical software packages.

WileyPLUS. This powerful online tool provides a completely integrated suite of teaching and learning resources in one easy-to-use website. *WileyPLUS* offers an online assessment system with full gradebook capabilities and algorithmically generated skill building questions. This online teaching and learning environment also integrates the entire digital textbook. To view a demo of *WileyPLUS*, contact your local Wiley Sales Representative or visit: www.wiley.com/college/wileyplus.

ACKNOWLEDGMENTS

We thank Minitab (State College, Pa.) and the SAS Institute (Cary, N.C.) for permission to include commands and output from their software packages. A special thanks to K. T. Wu and Kam Tsui for many helpful suggestions and comments on earlier editions. We also thank all those who have contributed the data sets which enrich the presentation and all those who reviewed the previous editions. The following people gave their careful attention to this edition:

Hongshik Ahn, Stony Brook University
 Prasanta Basak, Penn State University Altoona
 Andrea Boito, Penn State University Altoona
 Patricia M. Buchanan, Penn State University
 Nural Chowdhury, University of Saskatchewan
 S. Abdul Fazal, California State University Stanislaus
 Christian K. Hansen, Eastern Washington University
 Susan Kay Herring, Sonoma State University
 Hui-Kuang Hsieh, University of Massachusetts Amherst
 Hira L. Koul, Michigan State University
 Melanie Martin, California State University Stanislaus
 Mark McKibben, Goucher College
 Charles H. Morgan, Jr., Lock Haven University of Pennsylvania
 Perpetua Lynne Nielsen, Brigham Young University
 Ashish Kumar Srivastava, St. Louis University
 James Stamey, Baylor University
 Masoud Tabatabai, Penn State University Harrisburg
 Jed W. Utsinger, Ohio University
 R. Patrick Vernon, Rhodes College

Roumen Vesselinov, University of South Carolina
Vladimir Vinogradov, Ohio University
A. G. Warrack, North Carolina A&T State University

Richard A. Johnson
Gouri K. Bhattacharyya

Contents

1 INTRODUCTION

1

- 1 What Is Statistics? 3
- 2 Statistics in Our Everyday Life 3
- 3 Statistics in Aid of Scientific Inquiry 5
- 4 Two Basic Concepts—Population and Sample 8
- 5 The Purposeful Collection of Data 14
- 6 Statistics in Context 15
- 7 Objectives of Statistics 17
- 8 Using Statistics Wisely 18
- 9 Key Ideas 18
- 10 Review Exercises 19

2 ORGANIZATION AND DESCRIPTION OF DATA

21

- 1 Introduction 23
- 2 Main Types of Data 23
- 3 Describing Data by Tables and Graphs 24
 - 3.1 Categorical Data 24
 - 3.2 Discrete Data 28
 - 3.3 Data on a Continuous Variable 29
- 4 Measures of Center 40
- 5 Measures of Variation 48
- 6 Checking the Stability of the Observations over Time 60
- 7 More on Graphics 64
- 8 Statistics in Context 66
- 9 Using Statistics Wisely 68
- 10 Key Ideas and Formulas 68
- 11 Technology 70
- 12 Review Exercises 73

3 DESCRIPTIVE STUDY OF BIVARIATE DATA 81

- 1 Introduction 83
- 2 Summarization of Bivariate Categorical Data 83
- 3 A Designed Experiment for Making a Comparison 88
- 4 Scatter Diagram of Bivariate Measurement Data 90
- 5 The Correlation Coefficient—A Measure of Linear Relation 93
- 6 Prediction of One Variable from Another (Linear Regression) 104
- 7 Using Statistics Wisely 109
- 8 Key Ideas and Formulas 109
- 9 Technology 110
- 10 Review Exercises 111

4 PROBABILITY 115

- 1 Introduction 117
- 2 Probability of an Event 118
- 3 Methods of Assigning Probability 124
 - 3.1 Equally Likely Elementary Outcomes—
The Uniform Probability Model 124
 - 3.2 Probability As the Long-Run Relative Frequency 126
- 4 Event Relations and Two Laws of Probability 132
- 5 Conditional Probability and Independence 141
- 6 Bayes' Theorem 140
- 7 Random Sampling from a Finite Population 155
- 8 Using Statistics Wisely 162
- 9 Key Ideas and Formulas 162
- 10 Technology 164
- 11 Review Exercises 165

5 PROBABILITY DISTRIBUTIONS 171

- 1 Introduction 173
- 2 Random Variables 173
- 3 Probability Distribution of a Discrete Random Variable 176
- 4 Expectation (Mean) and Standard Deviation
of a Probability Distribution 185
- 5 Successes and Failures—Bernoulli Trials 193
- 6 The Binomial Distribution 198
- 7 The Binomial Distribution in Context 208
- 8 Using Statistics Wisely 211
- 9 Key Ideas and Formulas 212
- 10 Technology 213
- 11 Review Exercises 215

6 THE NORMAL DISTRIBUTION 221

- 1 Probability Model for a Continuous Random Variable 223
- 2 The Normal Distribution—Its General Features 230
- 3 The Standard Normal Distribution 233
- 4 Probability Calculations with Normal Distributions 238
- 5 The Normal Approximation to the Binomial 242
- *6 Checking the Plausibility of a Normal Model 248
- *7 Transforming Observations to Attain Near Normality 251
- 8 Using Statistics Wisely 254
- 9 Key Ideas and Formulas 255
- 10 Technology 256
- 11 Review Exercises 257

**7 VARIATION IN REPEATED SAMPLES—
SAMPLING DISTRIBUTIONS 263**

- 1 Introduction 265
- 2 The Sampling Distribution of a Statistic 266
- 3 Distribution of the Sample Mean and the Central Limit Theorem 273
- 4 Statistics in Context 285
- 5 Using Statistics Wisely 289
- 6 Key Ideas and Formulas 289
- 7 Review Exercises 290
- 8 Class Projects 292
- 9 Computer Project 293

8 DRAWING INFERENCES FROM LARGE SAMPLES 295

- 1 Introduction 297
- 2 Point Estimation of a Population Mean 299
- 3 Confidence Interval for a Population Mean 305
- 4 Testing Hypotheses about a Population Mean 314
- 5 Inferences about a Population Proportion 329
- 6 Using Statistics Wisely 337
- 7 Key Ideas and Formulas 338
- 8 Technology 340
- 9 Review Exercises 343

9 SMALL-SAMPLE INFERENCES FOR NORMAL POPULATIONS

349

-
- 1 Introduction 351
 - 2 Student's t Distribution 351
 - 3 Inferences about μ —Small Sample Size 355
 - 3.1 Confidence Interval for μ 355
 - 3.2 Hypotheses Tests for μ 358
 - 4 Relationship between Tests and Confidence Intervals 363
 - 5 Inferences about the Standard Deviation σ
(The Chi-Square Distribution) 366
 - 6 Robustness of Inference Procedures 371
 - 7 Using Statistics Wisely 372
 - 8 Key Ideas and Formulas 373
 - 9 Technology 375
 - 10 Review Exercises 376

10 COMPARING TWO TREATMENTS

381

-
- 1 Introduction 383
 - 2 Independent Random Samples from Two Populations 386
 - 3 Large Samples Inference about Difference of Two Means 388
 - 4 Inferences from Small Samples: Normal Populations with
Equal Variances 394
 - 5 Inferences from Small Samples: Normal Populations with Unequal
Variances 400
 - 5.1 A Conservative t Test 400
 - 5.2 An Approximate t Test—Satterthwaite Correction 402
 - 6 Randomization and Its Role in Inference 407
 - 7 Matched Pairs Comparisons 409
 - 7.1 Inferences from a Large Number of Matched Pairs 412
 - 7.2 Inferences from a Small Number of Matched Pairs 413
 - 7.3 Randomization with Matched Pairs 416
 - 8 Choosing between Independent Samples and a Matched Pairs Sample 418
 - 9 Comparing Two Population Proportions 420
 - 10 Using Statistics Wisely 426
 - 11 Key Ideas and Formulas 427
 - 12 Technology 431
 - 13 Review Exercises 434

11 REGRESSION ANALYSIS—I Simple Linear Regression

439

-
- 1 Introduction 441
 - 2 Regression with a Single Predictor 443

- 3 A Straight-Line Regression Model 446
- 4 The Method of Least Squares 448
- 5 The Sampling Variability of the Least Squares Estimators—
Tools for Inference 456
- 6 Important Inference Problems 458
 - 6.1. Inference Concerning the Slope β_1 458
 - 6.2. Inference about the Intercept β_0 460
 - 6.3. Estimation of the Mean Response for a Specified x Value 460
 - 6.4. Prediction of a Single Response for a Specified x Value 463
- 7 The Strength of a Linear Relation 471
- 8 Remarks about the Straight Line Model Assumptions 476
- 9 Using Statistics Wisely 476
- 10 Key Ideas and Formulas 477
- 11 Technology 480
- 12 Review Exercises 481

12 REGRESSION ANALYSIS—II

Multiple Linear Regression and Other Topics

485

- 1 Introduction 487
- 2 Nonlinear Relations and Linearizing Transformations 487
- 3 Multiple Linear Regression 491
- 4 Residual Plots to Check the Adequacy of a Statistical Model 503
- 5 Using Statistics Wisely 507
- 6 Key Ideas and Formulas 507
- 7 Technology 508
- 8 Review Exercises 509

13 ANALYSIS OF CATEGORICAL DATA

513

- 1 Introduction 515
- 2 Pearson's χ^2 Test for Goodness of Fit 518
- 3 Contingency Table with One Margin Fixed
(Test of Homogeneity) 522
- 4 Contingency Table with Neither Margin Fixed (Test of Independence) 531
- 5 Using Statistics Wisely 537
- 6 Key Ideas and Formulas 537
- 7 Technology 539
- 8 Review Exercises 540

14 ANALYSIS OF VARIANCE (ANOVA)

543

- 1 Introduction 545
- 2 Comparison of Several Treatments—
The Completely Randomized Design 545

3	Population Model and Inferences for a Completely Randomized Design	553
4	Simultaneous Confidence Intervals	557
5	Graphical Diagnostics and Displays to Supplement ANOVA	561
6	Randomized Block Experiments for Comparing k Treatments	563
7	Using Statistics Wisely	571
8	Key Ideas and Formulas	572
9	Technology	573
10	Review Exercises	574

15 NONPARAMETRIC INFERENCE

577

1	Introduction	579
2	The Wilcoxon Rank-Sum Test for Comparing Two Treatments	579
3	Matched Pairs Comparisons	590
4	Measure of Correlation Based on Ranks	599
5	Concluding Remarks	603
6	Using Statistics Wisely	604
7	Key Ideas and Formulas	604
8	Technology	605
9	Review Exercises	605

APPENDIX A1	SUMMATION NOTATION	609
APPENDIX A2	RULES FOR COUNTING	614
APPENDIX A3	EXPECTATION AND STANDARD DEVIATION—PROPERTIES	617
APPENDIX A4	THE EXPECTED VALUE AND STANDARD DEVIATION OF \bar{X}	622

APPENDIX B	TABLES	624
	Table 1 Random Digits	624
	Table 2 Cumulative Binomial Probabilities	627
	Table 3 Standard Normal Probabilities	634
	Table 4 Percentage Points of t Distributions	636
	Table 5 Percentage Points of χ^2 Distributions	637
	Table 6 Percentage Points of $F(\nu_1, \nu_2)$ Distributions	638
	Table 7 Selected Tail Probabilities for the Null Distribution of Wilcoxon's Rank-Sum Statistic	640
	Table 8 Selected Tail Probabilities for the Null Distribution of Wilcoxon's Signed-Rank Statistic	645
DATA BANK		647
ANSWERS TO SELECTED ODD-NUMBERED EXERCISES		665
INDEX		681

